THE EVOLUTION OF "WHY?"

Daniel Dennett

I The descent of content: Which way is up?

I have recently completed the bracing – and long overdue – experience of reading *Making It Explicit*, and trying to superimpose my own views on it to see where the disagreements stand out. Mainly, there is agreement. Claim after claim, page after page, I find myself agreeing wholeheartedly with Brandom, and this is not just on matters about which philosophers in general agree. Brandom and I are on the same page about many issues that divide the profession. Moreover, on a few points where Brandom explicitly objects to positions I have maintained, I think he is, in the main, right, and I will acknowledge this in more detail below. What then is left to argue about? There is a fundamental difference of *direction* in our work. But I will try to show that our residual disagreements – which at first make us appear poles apart on some very central theses – *might* quite easily be nudged into consonance if Brandom can take on what I view as a few friendly amendments.

We both learned a lesson from Wilfrid Sellars that still hasn't sunk in with many of our colleagues. I quote, not for the first time, what I consider to be the pithiest expression of it in Sellars:

> My solution is that "'. . .' means - - -" is the core of a unique mode of discourse which is as distinct from the *description* and *explanation* of empirical fact, as is the language of *prescription* and *justification*.
> (Chisholm and Sellars, 1958, p. 527 – discussed briefly by me in 1987, p. 341)

The ineliminable, foundational normativity of all talk of meaning or intentionality was first insisted upon by Sellars, and Brandom's version of the reason for this is comprehensive and detailed. Brandom chooses to adopt my "intentional stance" way of characterizing this unique mode of discourse, since the evaluative or normative presuppositions can be readily seen to be built right into the rules of that game. He then draws a distinction between *simple intentional systems* (what I call first-order intentional systems, entities whose behavior is readily interpretable by ascribing beliefs, desires, and other intentional states to them) and *interpreting*

intentional systems (what I call higher-order intentional systems, capable of ascribing intentional states to others and to themselves). Now which kind comes first?

The question is ambiguous, and once we sort it out, we'll have exposed a central int of discord between Brandom and me. Of course Brandom and I agree on the obvious evolutionary fact: we interpreting intentional systems are a later-evolved and much more sophisticated sort of agent – indeed, human beings with language are the only species that uncontroversially meets the requirements, in spite of much hopeful research with apes, dolphins, and other promising species.

Shouldn't the earlier, simpler variety be viewed as more fundamental? As Brandom says, "it is clear that there were nonlinguistic animals before there were linguistic ones, and the latter did not arise by magic." (p. 155) As he notes, Stalnaker and I follow this course, attempting to make sense of the "nonlinguistic" or "instrumental" belief of such animals as the fundamental notion. Brandom appreciates at least some of the grounds for the move – "this is a laudable aspiration and it may seem perverse to spurn it" (p. 155) – but he insists on adopting the "contrary order of explanation":

> the intentionality of nonlinguistic creatures is presented as dependent on, and in a specific sense derivative from, that of their linguistically qualified interpreters, who as a community exhibit a nonderivative, original intentionality. (p. 152)

This might seem to suggest, bizarrely, that nonlinguistic creatures had no intentionality until we humans came along to interpret them, at which point they acquired derivative intentionality, but of course that is not what Brandom means. What he means is that nonlinguistic creatures, long before human beings came along to interpret them, had a sort of cognitive prowess that is only *with strain* interpretable as a sort of intentionality, modeled on the sort of intentionality that we can make

se of only as a phenomenon occurring within a full-fledged linguistic commu-....y: "the understanding of intentionally contentful states that permits us to stretch the application of that notion and apply it in a second-class way to nonlinguistic animals (simple intentional systems)." (p. 151) The *order of explanation* is thus first to explain how linguistic communities carry on, relying on norms implicit in their social practices, and then use that account to provide a model, bristling with caveats and demurrals, for the more rudimentary phenomena exhibited by nonlinguistic animals and other simple intentional systems.

This is certainly not a position I have ever defended in detail, but I've made a few gestures in this direction, and I agree with most of it. Now that Brandom has done such a fine job making out the case for it, I am happy to subscribe to it – up to a point. It is not just that (of course) it takes language-using human beings to create the very terms and practices of ascription that render the intentional stance visible and available to science as a model for understanding non-language using agents. That would be like the fact that you could hardly speak of some chemical reaction as a "dance of the molecules" if there weren't any dancing people to serve

as exemplars. It is rather that the presuppositions of those terms and practices run deep, and until we understand them in their home territory, the deontic score-keeping game of asking for and giving reasons, we will not be able to understand the normativity that alone can give traction to a regress-stopper that we need. I take this to be the chief import of Brandom's title, *making it explicit*: until we language-users came along and made "it" explicit, *it* didn't really have intentionality – at least not the full-fledged intentionality we need to understand before we can extend the concept to our distant ancestors and other languageless varieties.

What is the problem? It can perhaps best be seen by following Brandom's vision of the contrast between his position and mine (we will later soften the contrast):

> Dennett's strategy of treating the normative significances of intentional states as instituted by the attitudes of interpreters does not by itself involve a commitment to reducing the normative to the nonnormative, insofar as it is the *proprieties* of attitudes that are invoked. That reductive commitment comes in later, in explaining these proprieties. Understanding those proprieties in terms of predictive success, as Dennett does (a strategy different from that to be pursued here) gives an objective basis to the norms governing the adoption of the intentional stance. It puts Dennett in a position to say that talk of the predictive utility of adopting that stance is just a way – indeed, the only one available to us – of specifying an important kind of objective pattern in behavior. (p. 58)

Brandom appreciates the need to ground his norms. They aren't just prejudices or fashions that have unaccountably caught on, and he can't resort to such anti-naturalist moves as *God commands them*. But still, he does not want to commit himself to any of the familiar kinds of naturalism either, which leaves him with a delicate task:

> the challenge: to maintain the stance stance[1] toward both simple and inter-preting intentional systems – that is, to acknowledge that the normative status of being such intentional systems is intelligible only by reference to the normative attitude of taking or treating something as such a system, that is interpreting it as one – while at the same time securing the distinc-tion between original and derivative intentionality – and so not allowing the notion of intentional normative status to collapse into that of the normative attitude of intentional interpretation. (p. 60)

Oversimplifying somewhat, the distinction Brandom sees between my naturalistic way of "collapsing" intentional normative status and his way of avoiding the regress can be captured by the supposed contrast between the violated norms of *faulty design* (my way) and the violated norms of a *social transgression* (his way). Roughly, it's the difference between being stupid and being naughty. At the same time, he must show that a violation of the social conceptual norms is not just a *faux pas*; it's a

mistake that really matters. I think it is this last obligation that in the end he cannot fulfill without falling back, eventually, on something like my way of dealing with the problem. As he notes, I attempt to meet this requirement with predictive utility – ! short and time flies and decisions must be made *now* by finite minds under cu...iderable pressure – but he finds my brand of pragmatism – in contrast with his own, actually quite similar, brand of pragmatism – to be unsatisfactory. What does he put in its place? Community.

II Original and derived intentionality

To understand how this ingenious move is supposed to work, consider the threatened regress: "Noises and marks on paper do not mean anything all by themselves." (p. 60) They don't have intrinsic intentionality but only derived intentionality. Derived from what? From the intentionality of *interpreting intentional systems* of course. And where do the interpreting intentional systems get their intentionality? Here is where our answers differ. Brandom says: *From each other!* "On this line, only communities, not individuals, can be interpreted as having original intentionality." (p. 61) And membership in a (linguistic) community while in some sense optional for us, carries with it a commitment to the conceptual norms established and constituted by that community.

I, in contrast, have claimed that *our own* intentionality is just as derived as that of our shopping lists and other meaningful artifacts, and derives from none other than Mother Nature, the process of evolution by natural selection. Consider a cartoonish dialogue about my version.

> We come upon a scrap of paper with the symbols "2 qt mk" on it.
> "What might it mean?"
> "Ask the agent who wrote it – the robot R1D1."
> R1D1 utters: "It's my shopping list and it means *buy two quarts of milk*."
> 'But where did R1D1 obtain *its* intentionality (assuming it has some)?"
> "From me, R1D1's designer. I made it to do my shopping for me, among other things."
> "And where did you get your intentionality?"
> "From *my designer*, Mother Nature!"
> "But what kind of purposes and intentions does 'she' have?"

None, in one sense, but plenty, in another: we can uncover the "free-floating rationales" of the design "decisions" that evolution has made, and thereby uncover the purposes and rationales of my organs and dispositions just as reverse engineering can uncover the purposes and rationales that account for the derived intentionality of my robot's organs and dispositions. The fact that evolution doesn't have purposes or intentions is not a bug but a feature: it permits us to end the regress of ulterior purposes with a whimper, not a bang. (The problem with God as the Intelligent Designer is that it forces the question of who designed God. The answer that nobody

had to design God is just as unsatisfactory as Searle's mystical – or at least mystifying – insistence that it's just a brute fact that human beings have original intentionality. That is why Searle's vision of original intentionality is ultimately incoherent; it begins with an unacceptably marvelous and unexplained prime mover. Brandom's account at least has the virtue of offering an explanation of how it is that the norms on which original intentionality depend play their roles. But he then has to face his own version of the question: whence – and why – came communities?)

As should be clear from the little dialogue above, my strategy has been to help myself to communicating (or "communicating") robots and their makers, and not make a big deal about the fact that this presupposes a community of communicators – the robot-makers – so that I can draw attention to the way in which the robot's intentionality (which I claim is well-nigh indistinguishable from ours in practice) depends on design considerations about which the robot-makers are the (corrigible) authorities. (They may have forgotten, or never understood, the good reasons why they made their robots the way they did.) In our own case, the parallel curiosity directs our attention to evolution. Just as we reverse-engineer the robot in order to understand its (derived) intentionality, so we reverse-engineer ourselves in order to understand the source of our own (derived!) intentionality. The only "original" intentionality anywhere is the mere *as-if* intentionality of the process of natural selection viewed from the intentional stance. Contrast this with Brandom's declaration: "Original, independent, or nonderivative intentionality is an exclusively linguistic affair." (p. 143)

Concentrating on the need to stop the regress of derived intentionality with a *diminuendo*, not a *crescendo*, I passed over – *glissando*, to continue the leitmotif – the issue of a *community* of communicating robot-makers as if it might be a local accident of history, rather than a constitutive requirement, but I now see no reason not to agree that Brandom is right about this – but not that the story ends where he ends it. I am willing, that is, to put in a major *fermata* at the place in the development where community comes in, and help myself to much of Brandom's book to provide an analysis of this. In fact, I spoke favorably – but I admit, only in passing – of just such a strategy some years ago, and in that passage foreshadowed both my current agreements and disagreements with Brandom, which I will at long last attempt to make explicit:

> [Michael] Friedman, discussing the current perplexity in cognitive psychology, suggests that the problem
>
> > is the direction of the reduction. Contemporary psychology tries to explain *individual* cognitive activity independently from *social* cognitive activity, and then tries to give a *micro* reduction of social cognitive activity – that is, the use of a public language – in terms of a prior theory of individual cognitive activity. The opposing suggestion is that we first look for a theory of social activity, and then try to give a *macro* reduction of individual cognitive activity – the activity of applying concepts, making judgments, and so forth – in terms of our prior social activity. (1981, p. 15–16)

With the idea of macro-reduction in psychology I largely agree, except that Friedman's identification of the macro level as explicitly social is only part of the story. The cognitive capacities of non-language-using animals (and Robinson Crusoes, if there are any) must also be accounted for, and not just in terms of an analogy with the practices of language users. The macro level *up* to which we should relate microprocesses in the brain in order to understand them as psychological is more broadly the level of organism–environment interaction, development, and evolution.

There is no way to capture the semantic properties of things (word tokens, diagrams, nerve impulses, brain states) by a micro-reduction. Semantic properties are not just relational but, you might say, super-relational, for the relation a particular vehicle of content, or token, must bear in order to have content is not just a relation it bears to other similar things (e.g., other tokens, or parts of tokens, or sets of tokens, or causes of tokens) but a relation between the token and the whole life – and counterfactual life – of the organism it 'serves' *and* that organism's requirements for survival *and* its evolutionary ancestry.

(Dennett, 1981, as reprinted in *The Intentional Stance*, 1987, p. 65)

III Are persons made of parrots?

Understanding can be understood, not as the turning on of a Cartesian light, but as practical mastery of a certain kind of inferentially articulated *doing*, responding differentially according to the circumstances of proper application of a concept, and distinguishing the proper inferential consequences of such application.

(Brandom, *Making It Explicit*, p. 120)

is is Brandom's brand of pragmatism, and I agree that you have to derive representation from inference, not vice versa. This is not recognized by many in philosophy though it has been common understanding in Artificial Intelligence for many years. William Woods' "What's in a Link" (1975) is the *locus classicus*. I myself have not always understood the implications of it, and I accept Brandom's use of me as a bad example of the formalist approach in "Intentional Systems," (1971, discussed by Brandom, pp. 99ff). Now that I see the work that "material inference" can do, I am ready to agree wholeheartedly with Brandom about this. I do not agree with him about the origins of norms, however.

Where do the norms come from and why? Let's consider Brandom's answer (in a typically hard-to-parse sentence, it must be said):

The objectivity of conceptual norms requires that any attitude of taking, treating, or assessing as correct an application of a concept in forming a belief or making a claim be coherently conceivable as mistaken, because of how things are with the objects the belief or claim is about. (p. 63)

53

Why? What is it about conceptual norms that imposes this requirement? Do communities just happen to have a correctness fetish (the original political correctness!)? The answer that I would want to give is shunned or ignored by Brandom: the social conceptual norms don't just happen; they are a requirement for a *working system* of communication. We as a species don't communicate just for the fun of it – though communication *is* a lot of fun, and that, too, is no accident. Communication has to pay for itself like every other complex adaptation, and this imposes design considerations among which are those that are met by this requirement.

Brandom accepts part of this: "it is clear that intentionality has a representational dimension and that to understand intentional contentfulness one must understand representation," but he shrinks from the idea that "a suitable notion of representation can be made available in advance of thinking about the correct use of linguistic expressions and the role of intentional states in making behavior intelligible." (p. 60)

It helps to understand the order of explanation here if we think of this as an engineering problem. Consider a so-called data-structure in a computer (as set up and maintained by a program – perhaps an airline reservation program). In what (strained) sense could it be *about* Chicago? If it serves to maintain, update, and provide information about Chicago to the functions it was designed to serve. It has to *work*: the numbers of the flights listed have to be flights into O'Hare Airport, and this fact has to be reliably maintained over time, and the data fields dealing with current weather have to track what the weather is in the Chicago area right now, and so on. If it does work, it is about Chicago in the only way that could matter here. Derived intentionality at its best: handsome is as handsome does.[2] The fact that these data-structures are typically written in such a way that they readily generate symbol strings that we human beings can read (e.g. they link to ASCII characters for "partly cloudy, wind SW at ten MPH") is an irrelevant feature. If a data-structure in a robot serves to maintain, update, and provide information about a wall in the robot's vicinity, and the robot needs to hug the walls to find its way out of a maze and uses the data-structure to guide itself (nonlinguistically, with even less sophistication than an insect), then that data-structure is about that wall in a fundamental sense that has nothing to do with linguistic expression.

And even though a robot – or an insect – is not a member of a linguistic community that chastises its members for violations of the community's norms, its data-structures are still in jeopardy of a sort of punishment: extinction for cause if they don't do what they are supposed to do. The robot's (intelligent) designer may be the critic who tells herself to go back to the drawing board and come up with a better system of data-structures, or she may have been intelligent enough to design the robot's information-processes to *correct themselves*, extinguishing the versions that don't work and replacing them with better versions – she's designed a learning robot, in other words. The same goes for the insect, of course. If the insect cannot learn to adjust its behavior when the world departs from the conditions in which its hardware evolved, it will succumb, *sphexishly* (Hofstadter, 1982; Hofstadter and Dennett, 1982; Dennett, 1984). The dividing line between "learning" by a lineage over evolutionary time, and learning, by individual trial-and-error, generate-and-test,

is not principled. Some mammals can walk at birth – they almost literally hit the ground running – and others need to learn how to control their legs. Evolution is supremely opportunistic in apportioning the R&D process between lineages and iduals.

Does the same verdict apply at the level of community instruction? Here, I think, is where Brandom thinks there is a significant watershed: there is all the difference in the world, he thinks, between norms maintained by communities of norm-protecting language-users and norms maintained by the relatively blind processes of natural selection or operant conditioning and its kin. But here he must confront his own version of the regress problem: even if we assume that normal adults appreciate the moves in the game – recognizing the speech acts that honor, flout, exploit, defend, abandon the norms of the linguistic community – how did they get themselves into this savvy condition? How, in short, does linguistic correction bootstrap itself into existence? John Haugeland (1998) confronts the same issue in his own version of Pittsburgh normativity, and "solves" it by speaking of an innate and pre-linguistic "censoriousness" that we humans are endowed with. A gift from God, or did it evolve, and if so, why? (See my review, Dennett, 1999.)

Why did "why" evolve? That is, why did the communicative community arise, and what sustains it? Brandom's pragmatist emphasis draws attention to the functional setting, explaining the *proximal* motivation for scorekeeping – the sanctions enforce the playing of the game – while ignoring the *distal* motivation: why does the game exist at all? What pays for this elaborate expenditure of energy?

Consider a dog that doesn't bark: the term "evolution" does not appear in the index of *Making It Explicit*. This is a measure of how resolutely Brandom has set his face against the "collapse" that I have proposed. *Community* is Brandom's skyhook (Dennett, 1995) and he can't have it. He knows this, but he prefers not to dwell on what it would take to secure community as his base of operations. Not surprisingly, this obliges him to reinvent the wheels of others without realizing it. Thus he sees misrepresentation or falsehood is parasitic on successful (and truthful) representation – "Purporting to represent is intelligible only as purporting to represent *successfully* or *correctly*" (p. 72) – but ignores the development of this argument in Dawkins and Krebs' (1978) classic work on the evolution of animal communication. Another term missing from the index is "Millikan", but Brandom speaks, for instance, of "some consumer or target" of a representation, a theme Millikan has developed in depth, treating *subpersonal* consumers of (subpersonal) representations (see Millikan, 2000). The subpersonal machinery of cognition is another arena that Brandom refuses to enter, though many of the themes he explores are well-explored by Millikan and others working that territory. My point is not a procedural criticism – he should have cited this work – but a substantive one: by ignoring it, he creates the illusion (for himself and his readers) that his community-*based* account of meaning and representation can be an autonomous *alternative* to a naturalistic theory of the same topics. The themes he discusses so insightfully are not just accessible to naturalism; they are (already) explicable, to a significant degree, in naturalistic terms.

I think it is instructive to trace the roughly parallel paths of some of Brandom's observations and their counterparts in naturalistic accounts. Starting in the physical world (from the physical stance, in my terms), Brandom notes that "iron doesn't conceive its world as wet when it responds by rusting." (p. 87) Why not? Because, as he says, "a normative dimension is required, which can underwrite a distinction between correct and incorrect application of concepts." He has just quoted Hegel's wonderful (and as he notes, naturalistic) account of animal desire – "an animal classifies some particular as food when it 'falls to without further ado and eats it up'" (p. 86) – but he neglects to note that this "responsive disposition" in the animal, unlike the disposition of the iron, is responsive to a normative dimension. Animals – and plants, too, of course – exploit basic, undesigned physical dispositions like the disposition of iron to rust in the transducers and other hardware that they are endowed with, *so that in the main what they eat is edible and nutritious.* These innate endowments provide plants and animals with sources of information – and, abnormally (in Millikan's sense), misinformation – which subpersonal consumers utilize in the processes of governance that keep life going. But in Brandom's terms, none of the transactions within the nervous system contain *explicit* meanings because although the transactors may be loosely, metaphorically, modeled as *homunculi* – cognitive agents of one sort or another – there isn't enough of a *social* system in place. There may be, as Minsky (1985) has put it, a society of mind, but there is no community!

Consider the parafoveal change-detectors or anomaly-detectors in the vision system that compete with each other to determine where our eyes dart next. "Look over here! Something interesting is happening in my sector," they all seem to shout, and the one that wins succeeds in attracting a saccade – the high-resolution foveas of the eyes jump to that point in the visual world for upgraded processing. But these signals are not speech acts; they are neither explicitly imperative ("Look over here!") nor declarative ("Something interesting is happening in my sector") but rather too primitive to count. One might even say they are degenerate, from Brandom's point of view, like Wittgenstein's "Slab!" language game, which Brandom says is not an instance of a verbal, but "merely vocal", language game (p. 172). (Cf. also the pioneering analysis of this issue by Jonathan Bennett, in his discussion of bee-dance "communication" in *Rationality*, 1964.)

In what to me is a key example, Brandom describes a red-measuring instrument hooked up to a tape recorder that appropriately emits "That's red" when a red thing is present. He compares it to a human observer who has been asked (hired?) to do the same thing. There is a world of difference, as he insists, and then he supplies a nice intermediate case: a parrot trained to do the same job (p. 88). In spite of being alive, the parrot is more like the instrument, Brandom insists, than the human observer, because the parrot lacks "a kind of understanding" – the kind of understanding Sellars rightly insisted is the key to content. Brandom says:

> Insofar as the repeatable response is not, for the parrot, caught up in practical proprieties of inference and justification, and so of the making of further judgments, it is not a conceptual or cognitive matter at all. (p. 89)

Not at all? This needs to be adjusted. The parrot is *some* kind of cognitive agent, and even if it doesn't *explicitly* engage in inference and justification, its cognitive processes include quality-control measures that track such explicit processes to a signifi- extent. The contrast Brandom wants to draw between persons and parrots is no. as big a gulf as he makes out, but it is real, and it is important *precisely because it helps us see a way out of our regress problem*: in order for there to be such large agents – persons – capable of Sellarsian understanding, they need to be composed of smaller, stupider agents with much less understanding – not so much subpersonal *homunculi* as *psittaculi!*

Organisms are correctly seen as multi-cellular communities sharing, for the most part, a common fate (they're in the same boat). So evolution can be expected to favor cooperative arrangements in general. Your eyes may, on occasion, deceive you – but not on purpose! (See Sterelny, 2003.) Running is sure to be a coordinated activity of the limbs, not a battle for supremacy. Nevertheless, there are bound to be occasions when subsystems work at cross purposes, even in the best-ordered communities of cells, and these will in general be resolved in the slow, old-fashioned way: by the extinction of those lineages in which these conflicts arise most frequently. The result is control systems that get along quite well *without* any internal self-monitoring. The ant colony has no boss, and no virtual boss either, and gets along swimmingly with distributed control that so far as we can tell does not engage or need to engage in high-level self-monitoring. According to the ethologist and roboticist David McFarland (1989), "Communication is the only behavior that requires an organism to self-monitor its own control system." Organisms can very effectively control themselves by a collection of competing but "myopic" task-controllers that can interrupt each other when their conditions ("hunger" or need, sensed opportunity, built-in priority ranking, ...) outweigh the conditions of the currently active task-controller. Goals are represented only tacitly, in the feedback loops that guide each task-controller, but without any global or higher-level representation. (One r 't think of such a task-controller as "uncommented code" – it works, but there i. thing anywhere in it that can be read off about what it does or why or how it does it.) Evolution will tend to optimize the interrupt dynamics of these modules, and nobody's the wiser. That is, *there doesn't have to be anybody home to be wiser!*

But communication, McFarland thinks, is a behavioral innovation that changes that. Communication requires a central clearing house of sorts *in order to buffer the organism from revealing too much about its current state* to competitive organisms. In order to understand the evolution of communication, as Dawkins and Krebs (1978) showed, we need to see it as *manipulation* rather than as purely cooperative behavior. The organism that has no poker face, that communicates state directly to all hearers, is a sitting duck, and will soon be extinct. What must evolve instead is a communication-control buffer that creates (1) opportunities for guided deception, and coincidentally (2) opportunities for self-deception (Trivers, 1985), by creating, for the first time in the evolution of nervous systems, explicit (and more "globally" accessible) *representations* of its current state, representations that are detachable from the tasks they represent, so that deceptive behaviors can be formulated and

57

controlled.[3] This in turn opens up structure that can be utilized in taking the step, described in detail by Gary Drescher (1991), from simple *situation-action machines* to *choice machines*, the step I describe as the evolutionary transition from Skinnerian to Popperian creatures (Dennett, 1995). (The previous two paragraphs are drawn from Dennett, 2007.)

Supposing that we can analyze whole persons with all their Sellarsian understanding as composed of less comprehending interactors is not just a convenience to the theorist, for a reason that Brandom spells out: individuation by content is inescapably holistic because it involves both premises and conclusions (p. 90). We can't just reach into the *snow is white* bin and pull out a *snow-is-white* to install in our cognitive system's belief box! There is no such thing as a *snow-is-white* independently of the huge system in which it must reside. So items individuated by content *cannot* be building blocks, independently identified. But we do need building blocks, so we can conclude that our blocks must be individuated by something "less" than semantic content. We need Janus-faced things that look, from some angles, rather like propositional contents, and from other angles like scraps of machinery. Bring on the parrotings, understood derivatively as if they were (almost) speech acts. And expect that Brandom's analyses will prove to be valuable tools for measuring the ways in which subpersonal agencies fall short.

For instance, at what level(s) does *deontic status* of one or another attenuated sort arise? What is the difference between the guardian role of the immune system, for instance, and the "ticket taker or doorman" Brandom usefully imagines? Could there be *a sort* of social scorekeeping among the semi-independent, cooperating/competing subsystems in the nervous system? Suppose we were to take Brandom's nicely elaborated model of interpersonal communication and attempt to apply it dead literally to the communication between specialist agencies in the brain. What features would drop out and why? According to Brandom, it is only the "practical inferential proprieties" tracked in the scorekeeping that "make noises and marks mean what they mean," but aren't there somewhat attenuated – *de-socialized*, one might say – inferential processes that can make brain-signals mean what they mean? Whatever the answer, it will be a valuable contribution to the theoretical understanding of the ways in which nervous systems work their unmagical magic.

IV Beliefs and opinions: arresting Brandom's flight from naturalism

I don't think there has to be a disagreement between Brandom and me at this point. Assuming that his neglect of evolutionary considerations isn't some sort of closet creationism (not that that is unknown among eminent philosophers dealing with these topics), he ought to be able to take on board most if not all of my friendly amendments. He still gets to maintain the ineliminably social or communitarian grounding of meaning, and thereby explain the second-class or derived application of the constitutive concepts to animals, to human parts, and, of course, to evolution itself. I agree with him about the conceptual dependence of all treatments of

subpersonal parrotings as derivative from, dependent on, our prior understanding of personal-level moves in the deontic scorekeeping game of inference, and I *even* agree that "Beliefs are essentially the sort of thing that can be expressed by making assertion." (p.153) How can I agree with this? By acceding, in a discussion of his book, to his specified use of the term "belief":

> There clearly is a sense in which nonlinguistic animals can be said to have beliefs. But the sense of belief that Sellars, Dummett, and Davidson are interested in (and that is the subject of this work) is one in which beliefs can be attributed only to language users. (p. 155)

So a belief for Brandom is very much what I call an *opinion* in *Brainstorms* (1978): the state of acceptance of a sentence collected (by a linguistic creature, naturally) as true, contrasting such a sophisticated state with the "lower brutish state" of belief (p. 305). This makes commentary awkward since I use "belief" and "opinion" as contrasting terms, and Brandom uses "belief" more or less as I use "opinion" and contrasts beliefs so-characterized with "the intentional states of nonlinguistic creatures," a category for which he doesn't have a shorter term. I'm quite comfortable with Brandom's insistence that such a language-involving acceptance state is not just the paradigm of a contentful state but that without which we couldn't really make sense of attributing *specific* contents in more *attenuated* cases. So I agree that "what the frog's eye tells the frog's brain" and what the dog or the chess-playing computer *believes* (in my sense) has to be understood, as he insists, as a "stretching" usage that takes us out of the home cases into a world of derivative cases. "Where intentional explanations are offered of the behavior of nonlinguistic creatures (those that are not understood as interpreters of others), the reasons are offered, the assertions are made, by the interpreter of a simple intentional system, who seeks to make its behavior intelligible by treating it *as if* it could act according to reasons it offers lf." (p. 171) Its reasons are what I call "free-floating rationales" that it does not resent to itself, and has no need of representing to itself. When, then, does the need arise? This is one of the important – foundational – questions left unasked by Brandom's resolutely upward bound account. And in the lacuna thus created he inserts a dubious *apologia*.

He assumes that "suitable social creatures can learn to distinguish in their practice between performances that are treated as correct by their fellows (itself a responsive discrimination) and those that are not" and he is careful to avoid circularity by insisting that "no appeal will be made to instrumental rationality on the part of fledgling linguistic practitioners" (p.155). In other words, since he cannot help himself to the intentional stance yet, he has to follow a fundamentally Skinnerian line, behaviorism, replacing a *rational* or *cognitive* account of this learning process with a suitably innocent *conditioning* account. And this decision, which he thinks he needs to take to avoid circularity or vicious regress, creates a problem for him, parallel to the question that has bedeviled Fodor, Dretske, and others under the guise of the (ill-posed) *disjunction problem*: what can the theorist appeal to in

either an individual's history of learning by trial-and-error or a lineage's history of evolutionary trial-and-error to distinguish errors from mere extensions of the category being discriminated? (See my "Evolution, Error and Intentionality," in *The Intentional Stance*, 1987.) Brandom recognizes the problem: "How is it possible for our use of an expression to confer on it a content that settles that we might all be wrong about how it is correctly used, at least at some times?" (p. 137) His version has the advantage that he gets to appeal to community practices and community agreement about how the norms are to be applied, and he sees that his answer, like a good answer to the disjunction problem, has to find, as he says, some way of privileging a disposition – except that he doesn't actually say how he is going to do this.

He excuses himself from this task by retracting into a very modest goal: "explaining what the trick consists in, what would count as doing it, rather than how it is done by creatures wired up as we are." (p. 155) In other words, he wants to give the "specs" without a word on the implementation, the (reverse) engineering. This is a remarkable truncation, given not just the ambition but the gratifying success of his own reverse engineering of the game of asking and giving reasons. He blandly assures the reader that "the abilities attributed to linguistic practitioners are not magical, mysterious or extraordinary" (p. 156) but we are entitled to ask an emperor's new clothes question: "Why not?"

"There were no commitments before people started treating each other as committed" – a phrase of Brandom's that echoes Hobbes' celebrated attempt to reconstruct the origins of morality out of the state of nature, and Brandom can't eschew discussing this transition by simply assuring us that it wasn't magic. For instance, why haven't other species evolved something like promising, for instance? It's obviously *good for something*. Hobbes presupposed language among his people in the state of nature, and Brandom's claim must be that the deontic status of language – with its sanctioned practices and hence implicit concepts of authority, commitment and entitlement – already contains something with a variety of normativity missing in the state of (nonhuman) nature. (See p. 175.) Whether the commitments of language are properly the *ancestors* of moral commitments or merely had to co-evolve with them is an interesting question, not raised by Brandom.

This is, I think, an unacceptable flight from naturalism, not just because it refuses to address an entirely appropriate question but because by doing so it actually distorts the analysis at the higher level. A deeper account of communication would be a better account. When organisms take on the new behavior of signaling (as contrasted with simply willy-nilly providing information about their state by being perceptible by their audience) the opportunity for strategic deception arises, and plays an important role in fixing precisely the sort of dependencies Brandom himself is arguing for. He asserts, convincingly, that "These pragmatic inferential practices form a shell around the more basic semantic ones, which they presuppose" (p. 159) but both also presuppose the underlying emergence of communication as a behavioral opportunity, with its own costs and benefits. Credible signaling must be hard-to-fake signaling, lest inflation set in and destroy the communication channel with noisy fakery. The panglossian (or polyannian) presumption of cooperation must not be built in.

These are quite uncontroversial themes in the considerable naturalistic literature on the evolution of communication. Adding them to his perspective would not commit Brandom to any doctrinaire *scientism* or *evolutionism* that might seem to (:en the philosophical austerity of his project, and it would go some way to removing the suspicious residue of magic that clings to any account that purports to ground meaning in community and then, having done that, declares victory.[4]

Notes

1   For Brandom it is always the intentional stance, never the design stance that is under discussion.
2   Dummett's emphasis on "harmony" between the introduction rules (the circumstances) and the elimination rules (the consequences) is a good way of bringing out the "engineering" aspect, the need for the system to *work*. Dummett's excellent example is the term "Boche" for Germans, which has, as Brandom stresses, *inappropriate* and *indefensible* content (p.125ff).
3   Millikan, 2000, says "When it is a natural purpose that is represented, this correspond-ence relation correlates the representation with a state of affairs that it is its proper function to guide a cooperating mechanism to bring about." [p86] This is rather like the Chicago-representation in the computer: there is in effect a variable that can take the value Chicago or other, contrasting values, such as Boston. This is a *kind* of explicit representation, but not either McFarland's or Brandom's kind. (See my "Styles of Mental Representation," in *The Intentional Stance*.)
4   Thanks to Richard Griffin, Nicholas Humphrey and Roger Scruton for valuable edito-rial suggestions about an earlier draft.

References

Bennett, Jonathan (1964) *Rationality* (London: Routledge & Kegan Paul).

Chisholm, R. and Sellars, W. (1958) "The Chisholm–Sellars Correspondence on Intentionality," in H. Feigl, M. Scriven and G. Maxwell, eds., *Minnesota Studies in the Philosophy of Science, Vol. II* (Minneapolis: University of Minnesota Press), pp. 529–39.

ins, R., and Krebs, J. R., (1978) "Animal Signals: Information or Manipulation?" in R. Krebs and N. B. Davies, eds., *Behavioral Ecology An Evolutionary Approach* (Oxford: Blackwell), pp. 282–309.

Dennett, Daniel C. (1971) "Intentional Systems." *Journal of Philosophy* 68 (February), pp. 87–106.

— (1978) *Brainstorms* (Cambridge, MA: MIT Press).

— (1981) "Three Kinds of Intentional Psychology," in R. Healey, ed., *Reduction, Time and Reality* (Cambridge: Cambridge University Press), pp. 37–61, reprinted in Dennett, 1987.

— (1984) *Elbow Room: The Varieties of Free Will Worth Wanting* (Cambridge, MA: MIT Press).

— (1987) *The Intentional Stance* (Cambridge, MA: MIT Press).

— (1995) *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (New York: Simon & Schuster, London: Penguin).

— (1999) Review of John Haugeland: *Having Thought: Essays in the Metaphysics of Mind*, in *Journal of Philosophy* 96/8, pp. 430–5.

— (2007) "My Body has a Mind of its Own," in D. Ross, D. Spurrett, H. Kincaid, G.L. Stephens (eds), *Distributed Cognition and the Will: Individual Volition and Social Context* (Cambridge, MA: MIT Press, A Bradford Book), pp. 93–100.

Drescher, Gary (1991) *Made-Up Minds: A Constructivist Approach to Artificial Intelligence* (Cambridge, MA: MIT Press).

Haugeland, John (1998) *Having Thought: Essays in the Metaphysics of Mind* (Cambridge, MA: Harvard University Press).

Hofstadter, D. R. (1982) "Can Creativity be Mechanized?" *Scientific American* 247, pp. 20-9.

Hofstadter, D.R. and Dennett, D. (1982) *The Mind's I* (New York: Basic Books).

McFarland, David (1989) "Goals, No-Goals and Own Goals," in Alan Montefiore and Denis Noble (eds), *Goals, No-Goals and Own Goals: A Debate on Goal-Directed and Intentional Behaviour* (London: Unwin Hyman Ltd).

Millikan, Ruth (2000) "Naturalizing Intentionality," in Bernard Elevitch (ed.), *Philosophy of Mind, Proceedings of the Twentieth World Congress of Philosophy*, vol. 9 (Philosophy Documentation Center, 2000), pp. 83-90.

Minsky, Marvin (1985) *The Society of Mind* (New York: Simon & Schuster).

Sterelny, Kim (2003) *Thought in a Hostile World: The Evolution of Human Cognition* (Oxford: Blackwell).

Trivers, Robert (1985) *Social Evolution* (Menlo Park, CA: Benjamin/Cummings).

Woods, William (1975) "What's in a Link: Foundations for Semantic Networks," in D. Bobrow and A. Collins (eds), *Representation and Understanding: Studies in Cognitive Science* (New York: Academic Press). Reprinted in R. Brachman and H. Levesque (eds), *Readings in Knowledge Representation* (San Mateo, CA: Morgan Kaufmann, 1985). Also reprinted in Allan Collins and Edward E. Smith (eds), *Readings in Cognitive Science* (San Mateo, CA: Morgan Kaufmann, 1988).